

Державний вищий навчальний заклад  
«Донецький національний технічний університет»  
Кафедра прикладної математики та інформатики

«ЗАТВЕРДЖУЮ»

Перший проректор

\_\_\_\_\_ Леонід Бачурін

«\_\_\_\_\_» \_\_\_\_\_ 2024 р.

## РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

ВБ 2.1 ІНТЕЛЕКТУАЛЬНА ОБРОБКА WEB ДАНИХ

(шифр і назва навчальної дисципліни)

Рівень освіти: другий (магістерський)

Спеціальність: 122 Комп'ютерні науки  
(шифр і назва спеціальності (тей))

Освітня програма: Комп'ютерні науки  
(назва освітньої програми)

Мова навчання: українська

Робоча програма навчальної дисципліни «Інтелектуальна обробка Web даних» для здобувачів вищої освіти за спеціальністю 122 Комп'ютерні науки  
«25» грудня 2023 р. – 8 с.

Розробник: Ярош І.В., старший викладач кафедри прикладної математики та інформатики.

Робоча програма затверджена на засіданні кафедри прикладної математики та інформатики

Протокол № 13 від «27» грудня 2023 р.

Зав. кафедри \_\_\_\_\_ ( Маслова Н.О. )

«27» грудня 2023 р.

Схвалено науково-методичною комісією з галузі знань 12 «Інформаційні технології»

Протокол № 1 від «15» січня 2024 р.

Голова \_\_\_\_\_ ( Башков Є.О. )

## 1. Загальна інформація

Форма навчання	Денна	Заочна
Статус	Вибіркова	
Обсяг в кредитах ЄКТС	6	
Обсяг в годинах за навчальним планом, разом: в тому числі:	180	
лекції:	32	8
практичні заняття:	32	8
лабораторні заняття:	—	—
семінари:	—	—
самостійна робота:	116	164
Форма підсумкового контролю	Екзамен	
Дисципліну викладають	Викладач Ярош Ірина Вікторівна, <a href="https://donntu.edu.ua/kitaer/pmi">https://donntu.edu.ua/kitaer/pmi</a> , <a href="mailto:iryna.yarosh@donntu.edu.ua">iryna.yarosh@donntu.edu.ua</a> , <a href="https://wiki.donntu.edu.ua/view/Ярош_Ірина_Вікторівна">https://wiki.donntu.edu.ua/view/Ярош_Ірина_Вікторівна</a>	

**Передумови для вивчення дисципліни:** успішному вивченню дисципліни «Інтелектуальна обробка Web даних» сприяє попереднє опанування такими дисциплінами, як «Інтелектуальний аналіз даних», «Основи web-програмування», «Основи організації баз даних та знань», «Методи та системи штучного інтелекту», «Сучасні технології програмування», «Технології створення веб-застосунків», «Методи та засоби моделювання і представлення даних та систем», «Методи та засоби обробки текстової інформації».

## 2. Мета вивчення навчальної дисципліни «Інтелектуальна обробка Web даних»

**Метою** вивчення навчальної дисципліни є формування у студентів системи знань в області аналізу інформації, пов'язаної з Web-простором і мережею Internet, і її обробки за допомогою засобів інтелектуального аналізу даних, а також вмінь, навичок вибору та застосування інструментів Web Mining для оволодіння навичками для видобутку Інтернет-даних.

### Компетентності:

- здатність до абстрактного мислення, аналізу та синтезу;
- здатність застосовувати знання у практичних ситуаціях;
- здатність вчитися й оволодівати сучасними знаннями;
- здатність збирати і аналізувати дані (включно з великими), для забезпечення якості прийняття проєктних рішень;

- здатність застосовувати існуючі і розробляти нові алгоритми розв’язування задач у галузі комп’ютерних наук;
- здатність використовувати системні та базові знання зі штучного інтелекту, що включають в себе штучні нейронні мережі, розпізнавання образів та машинне навчання, експертні системи, нечітка логіка, агентні системи, розуміння природо-мовних текстів, автоматичне реферування;
- здатність застосовувати знання структури, функціонування, технічного та програмного забезпечення інтелектуальних та веборієнтованих комп’ютерних систем.

#### **Програмні результати навчання:**

- спеціалізовані концептуальні знання, що включають сучасні наукові здобутки у сфері комп’ютерних наук і є основою для оригінального мислення та проведення досліджень, критичне осмислення проблем у сфері комп’ютерних наук та на межі галузей знань;
- спеціалізовані уміння/навички розв’язання проблем комп’ютерних наук, необхідні для проведення досліджень та/або провадження інноваційної діяльності з метою розвитку нових знань та процедур;
- управління робочими процесами у сфері інформаційних технологій, які є складними, непередбачуваними та потребують нових стратегічних підходів;
- розробка математичних моделей та методів аналізу даних (включно з великими);
- розробка алгоритмічного та програмного забезпечення для аналізу даних (включно з великими);
- виявлення потреб потенційних замовників щодо автоматизації обробки інформації;
- виявлення та усунення проблемних ситуацій в процесі експлуатації програмного забезпечення, формулювання завдання для його модифікації або реінжинірингу.

### **3. Очікувані результати навчання**

Результатами опанування дисципліни «Інтелектуальна обробка Web даних» є наявність у студентів базових знань із основ, сучасної методології і володіння особливостями виконання інтелектуальної обробки вебданих, отриманих в мережі Internet; засвоєні теоретичні основи інформаційних систем, орієнтованих на застосування стандартів Data Mining в Internet (Web Mining); набуті уміння роботи з системами Web Mining різного призначення і різної проблемної орієнтації; отримані практичні навички використання та адаптації деяких найбільш відомих сервісів для Web Data Mining.

Внаслідок вивчення матеріалу навчальної дисципліни студент повинен знати: сутність і призначення технології Web Mining, її складові етапи та особливості їх виконання, існуючі категорії інтелектуального аналізу даних мережі Інтернет, основи застосування методів інтелектуального аналізу даних для вилучення знань із вебконтенту (вебдокументів і їх описів, доступних в Інтернеті), для видобутку та виявлення вебструктур, для аналізу використання вебресурсів; та вміти: досліджувати та використовувати вебконтент, аналізувати,

цілеспрямовано шукати та вибирати необхідні для вирішення фахових наукових і прикладних задач інформаційно-довідкові та науково-технічні ресурси і джерела знань з урахуванням сучасних досягнень методології Data Mining в WWW, оцінювати складові ефективності Web Mining, використовувати сервіси та бібліотеки для опрацювання даних мережі Internet і роботи з вебаналітикою.

#### 4. Засоби діагностики результатів навчання

Засобами оцінювання та методами демонстрування результатів навчання при опануванні дисципліни «Інтелектуальна обробка Web даних» є:

- екзамен;
- завдання індивідуальної роботи;
- завдання практичних робіт;
- презентації і обговорення результатів виконаних завдань та досліджень.

#### 5. Критерії оцінювання результатів навчання

Критерії оцінювання мають формулювати порядок оцінювання під час поточного контролю (за результатами практичних робіт і виконання індивідуальних завдань) та підсумкового контролю.

ІР1	ІР2	ІР3	ІР	Поточний контроль	Екзамен	Максимальний бал
10	12	12	6	40	60	100
6	7,2	7,2	3,6	24		

Примітка: 1) ІР1-ІР3 – практичні роботи, ІР – індивідуальна робота (розрахункова робота); 2) у числівнику максимальний бал – при своєчасному та правильному виконанні, у знаменнику – мінімальний (при правильному, але несвоечасному виконанні); 3) в оцінку поточного контролю з виконання практичних робіт включено контрольні та поточні опитування

Відповідність між шкалами встановлюється наступним чином:

Оцінка	
За 100-бальною шкалою	Для екзамену
90-100	відмінно
74-89	добре
60-73	задовільно
0-59	незадовільно

#### 6. Програма навчальної дисципліни

##### 6.1. Основні теми дисципліни

**Тема 1.** Розвиток, призначення та основи інтелектуального аналізу даних (ІАД, Data Mining).

**Тема 2.** ІАД в мережі Internet (Web Mining): сутність, особливості, проблематика, етапність, категорії.

**Тема 3.** Інструментарій для забезпечення Web Mining.

**Тема 4.** Виявлення структурної інформації в мережі Internet (Web Structure Mining).

**Тема 5.** Вилучення знань із вебконтенту (Web Content Mining).

**Тема 6.** Аналіз використання ресурсів мережі Internet (Web Usage Mining).

### 6.2. Теми практичних занять

№ з/п	Назва теми	Кількість годин:	
		д.ф.н.	з.ф.н.
1	Сутність і особливості інтелектуального аналізу даних мережі Internet (Основи Web Mining).	2	1
2	Інструментарій для вилучення та інтелектуального аналізу даних мережі Internet.	6	1
3	Виявлення структурної інформації в мережі Internet (Web Structure Mining).	8	2
4	Вилучення знань із вебконтенту (Web Content Mining).	8	2
5	Аналіз використання ресурсів мережі Internet (Web Usage Mining).	8	2
<b>Усього годин</b>		<b>32</b>	<b>8</b>

### 6.3. Теми лабораторних занять

Робочим навчальним планом лабораторні заняття не передбачені.

### 6.4. Самостійна робота

№ з/п	Назва теми	Кількість годин:	
		д.ф.н.	з.ф.н.
1	2	3	4
1	Тема 1. Розвиток, призначення та основи інтелектуального аналізу даних (ІАД, Data Mining).	6	12
2	Тема 2. ІАД в мережі Internet (Web Mining): сутність, особливості, проблематика, етапність, категорії.	14	20
3	Тема 3. Інструментарій для забезпечення Web Mining.	14	20
4	Тема 4. Вилучення знань із вебконтенту (Web Content Mining).	22	32
5	Тема 5. Виявлення структурної інформації в мережі Internet (Web Structure Mining).	26	36
6	Тема 6. Аналіз використання ресурсів мережі Internet (Web Usage Mining).	26	36
7	Виконання завдання індивідуальної роботи	8	8
<b>Усього годин</b>		<b>116</b>	<b>164</b>

## 6.5. Індивідуальне завдання

Робочим навчальним планом передбачено виконання індивідуальної роботи на тему «Дослідницький підхід до інструментів інтелектуального аналізу вебданих». Метою роботи є ознайомлення з існуючими інструментами, що використовується під час процесів видобутку даних і знань з мережі Internet, дослідження особливостей їх використання. Перелік пропонованих для обрання інструментів (засобів, сервісів і бібліотек) наведено в завданні індивідуальної роботи (Octoparse, Oracle Data Mining (ODM), Tableau, Scrapy, ParseHub, Diggernaut, Mozenda, Import.oi, Apify, Datahut і ін.). Варіант індивідуального завдання обирається студентом у відповідності до його порядкового номеру в переліку студентів академічної групи.

## 7. Література

### 7.1. Основна

1. Черняк О. І. Інтелектуальний аналіз даних : підручник / О. І. Черняк, П. В. Захарченко. К. : Знання, 2014. 599 с.
2. Аналіз даних та знань : навч. посібник / В. В. Литвин, В. В. Пасічник, Ю. В. Нікольський. Львів : «Магнолія 2006», 2021. 276 с.
3. Advanced Practical Approaches to Web Mining Techniques and Application / Obaid, Ahmed J., Polkowski, Zdzislaw, Bhushan, Bharat. IGI Global, 2022. 357 p.
4. Web Mining and it's application / Mr. V. Kathirvel. SK Research Group of Companies, 2023. 210 p.
5. Web Searching and Mining (Cognitive Intelligence and Robotics) / Debajyoti Mukhopadhyay. Springer, 2019. 276 p.

### 7.2. Додаткова (допоміжна)

1. Data Mining : пошук знань в даних / А. Я. Гладун, Ю. В. Рогушина. К. : ТОВ «ВД «АДЕФ-Україна», 2016. 452 с.
2. Web Content Mining aus dem Hidden Web : Ein methodischer Web-Mining-Ansatz für die automatisierte Extraktion von Informationen aus dynamischen Webseiten / Muhammad Asif Naeem. Verlag Unser Wissen, 2021. 88 p.
3. Web Usage Mining Techniques and Applications Across Industries (Advances in Data Mining and Database Management) / A. V. Senthil Kumar. Information Science Reference, 2016. 450 p.
4. Web Usage Mining of Hidden Knowledge Using Server Log Files / Wajih Abdul Ghani. LAP LAMBERT Academic Publishing, 2017. 152 p.
5. Web Mining : A Synergic Approach Resorting to Classifications and Clustering / V. S. Kumbhar, K. S. Oza, R. K. Kamat. River Publishers, 2022. 230 p.
6. Web Scraping with Python : Collecting More Data from the Modern Web / George Chang, Marcus Healey. O'Reilly Media, 2018. 478 p.
7. Michael Heydt. Python Web Scraping Cookbook : Over 90 proven recipes to get you scraping with Python, microservices, Docker, and AWS. Packt Publishing, 2018. 366 p.

### 7.3. Методична

1. Методичні вказівки до виконання практичних робіт з дисципліни «Інтелектуальна обробка Web даних» (у розробці).

### 8. Інформаційні ресурси

1. Web Mining : Applications and Techniques. URL: <http://pzs.dstu.dp.ua/DataMining/social/bibl/WebMining.pdf>.

2. Web Data Mining : Exploring Hyperlinks, Contents, and Usage Data. URL: [https://sirius.cs.put.poznan.pl/~inf89721/Seminarium/Web\\_Data\\_Mining\\_2nd\\_Edition\\_Exploring\\_Hyperlinks\\_Contents\\_and\\_Usage\\_Data.pdf](https://sirius.cs.put.poznan.pl/~inf89721/Seminarium/Web_Data_Mining_2nd_Edition_Exploring_Hyperlinks_Contents_and_Usage_Data.pdf).

Затверджене